Method for Video Enabled Electronic Commerce

BACKGROUND OF THE INVENTION

1.   The Field of the Invention

The present invention relates to electronic commerce, and more particularly to conducting electronic commerce by enabling creation of vision-enabled content.

2.   The Relevant Art

Activities such as advertising, entertainment and education are commonly conducted over a network such as the Internet. The creator of an activity conducts that activity by publishing content which then becomes available to users who are connected to the network and have the necessary program to receive and display that content, such as a web browser. For example, advertisements in the form of linked banners appear on a multitude of websites. Streaming audio and video as well as audio and video clips have become commonplace. Further, virtual classrooms and interactive learning materials are being used for long-distance learning.

Such activities, however, are constrained by the limitation of the technology being used to send, receive and navigate them. Users receiving content over a network currently interact with the content with input devices such as a mouse and keyboard. As a result, true interaction with the content must be left to the imagination. Users viewing an advertisement for shirts, for example, may be able to select different styles and colors of shirts, but would not be able to see himself or herself wearing the shirt. Since the user does not know how he or she looks in the shirt, the user is less likely to purchase the shirt from the advertiser and will more likely go to a store where the user may try the shirt on before purchase. Thus, the advertiser will probably lose the sale.

The problem is similar in entertainment. There are currently several products on the market which allow replay of downloaded audio and video. For example, Windows Media Player by Microsoft® Corporation and RealPlayer by RealNetworks,

Inc. allow an entertainment producer to transmit audio and video clips as well as streaming audio and video. Both of these products allow the user to interact with the content in the limited sense that the user is able to select the clips and streams and start and stop playback at will.

Unfortunately, these products are directed towards playback of content alone. Most users prefer to watch motion video on a television rather than over the Internet, typically because of the location and smaller size of the computer display. If the content is the same, there is little reason to watch it on the computer. There needs to be something that makes a user want to watch the content on the computer, such as a vision-based interaction between the user and the content.

Network gaming is a popular pastime for many people. While gaming technology has come far, gaming is still very impersonal in that the animated characters that represent each player bear only the likeness given it by the programmer and no resemblance to the actual player. Game play would be much more enjoyable if the animated characters of a game bore the likenesses of the associated players.

Adding to the impersonality of gameplay are game controllers. The realism of the game can often depend on how the player's commands are input into the computer. Movement of a user to make the animated character perform a similar movement is much more desirable than pushing a button to make a movement. Take, for example, a boxing game. A player would be much more likely to enjoy the game if the player could physically move his or her arm in a punching motion and see the animated character make a similar move in the game.

SUMMARY OF THE INVENTION

A method is provided for conducting commerce over a network via vision-enabled content. First, content is encoded to convert it into vision-enabled content. Payment is received for vision-enablement of the content. Also, a program to decode the vision-enabled content is provided. Finally, the vision-enabled content is sent to a user over a network. The program decodes the vision-enabled content and receives an image of the user. The vision-enabled content may include advertising content, entertainment content, and education content.

In one embodiment of the present invention, the program combines the image of the user with the vision-enabled content. The encoding allows a content publisher to distribute virtual content which can be received and interacted with by a user. For example, this would allow display of the user image interacting with a product or as part of entertainment content, such as an image of the user wearing a piece of clothing or along side a music star in a music video. This also allows a plurality of users to interact with each other, such as playing a game in which characters in the game bear resemblance to the users.

In another embodiment of the present invention, the program utilizes the image of the user to control the vision-enabled content. Controlling of the content includes not only selecting certain images based on the user image, but also controlling the way the content appears, such as using the person image to control the way a character moves through a game for example, with the game flow changing as a result of the actions of the character. In this way, a user is able to use movements to control the content being perceived by the user.

The encoding of the content may be performed via tools with payment being received in exchange for use of the tools. This allows a content provider to create its own vision-enabled content.

In one aspect of the present invention, payment may be received based on a number of users receiving the vision-enabled content. Alternatively, payment could be received based on a quantity, i.e., an amount, of vision-enabled content sent.

Payment may also be received from a content provider for storing the vision-enabled content.

In another aspect of the present invention, payment is received from the user. For instance, payment could be received from the user in exchange for the program. Optionally, an upgrade for the program can be offered. Payment could be received in exchange for the upgrade.

To personalize the content, an identity of a user may be recognized, such as from the person image, and the vision-enabled content can be selected based on the identity of the user. The user may also be associated with a group and the vision-enabled content selected based on the association of the user with the group.

As an option, body part recognition may be performed on the person image. This allows the user to assist in the selection of the content such as by performing a particular gesture. As an option, the content may be selected based on an interpretation of movement of the body part of the user.

The outputted content may include an interaction between the person image and the content, such as a portion of the person image appearing to interact with video images. As mentioned previously, body part recognition may be performed on the person image. In such case, the content may include an image of the body part of the user. The content may be output in real time via a data stream or sent in encapsulated form.

A background may be removed from the person image to assist in the recognition of a user in the person image. The background may also be removed to allow a portion of the person image to appear to interact with the content.

As an option, statistical data may be collected and used to create user profiles and informational databases. Optionally, payment may be received in exchange for access to the statistics.

These and other aspects and advantages of the present invention will become more apparent when the Description below is read in conjunction with the accompanying Drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be readily understood by the following detailed description in conjunction with the accompanying drawings.

Figure 1 is a diagram illustrating an interconnection between users, a content developer, a content publisher, and a technology provider in accordance with a business model of the present invention;

Figure 2 is a diagram illustrating components of a business model of the present invention;

Figure 3 illustrates an advertising model of the present invention corresponding to block A of Figure 2;

Figure 4 illustrates a process flow of the advertising model shown in Figure 3 in accordance with one embodiment of the present invention;

Figure 5A illustrates a process of the present invention associated with operation 412 of Figure 4 for personalizing content;

Figure 5B illustrates processes associated with operation 414 of Figure 4 in accordance with one embodiment of the present invention;

Figure 6 illustrates an entertainment/educational model of the present invention corresponding to blocks B and C of Figure 2; and

Figure 7 illustrates a process flow of the entertainment/educational model shown in Figure 6 in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is adapted for controlling content based on an image or series of images of a user. With reference to Figure 1, a user 100 connects to a content publisher's website 102 over a wide area network 104, e.g., the Internet, via a station, i.e., a computer, or other processing device such as a television. Vision-enabled content is sent to the user's computer in either encapsulated or streaming form over the wide area network 104 from the content publisher's website 102 where it is presented to the user via display or audio. An image or plurality of images of a user 100 are received and content based on the image of the user is selected from the vision-enabled content and displayed in such a way that the content appears to interact with the user 100, e.g., a portion of the image of the user 100 appears with the content, and/or movements of the user are recognized and used to control the content. More detail is provided below.

The content offered by the content publisher 102 may be created by a content developer 106 and sent to the content publisher. The tools, i.e., programs and hardware, necessary to encode the content into the vision-enabled format may be received from a technology provider 108. These programs and tools may be sent to either the content publisher or the content developer, or both.

Figure 2 is a diagram that illustrates various components a business model of the present invention. First, the content is encoded in a manner that converts it into vision-enabled content. The content may include streaming video, animated objects, web pages, games, advertising, educational applications, audio data, or anything else. As mentioned above, some or all of the tools 200 necessary to perform such encoding may be provided to a content provider, such as a content publisher or developer. This allows the content provider to create its own vision-enabled content. Payment would be received in exchange for use of these tools 200.

Alternatively, the technology provider may receive the content and perform the encoding of the content. Encoding fees 202 would be charged for performing the encoding. Once encoded, the content is sent to a publisher's website for dissemination.

The encoded content is sent to a user over a network. Preferably, the vision-enabled content is sent to the user's station via a data stream or in the form of an applet, where it is decoded by a program, e.g., a plug in 204. The data stream may be compressed. The plug in receives an image of the user from a camera 206 connected to the user's station. The applet may control the content based on the person image. Alternatively, the content may be controlled from the content provider's location. In either case, the user is allowed to interact with the content, as discussed below in more detail.

The plug in 204 and/or applet could be downloaded from the technology provider or content publisher. Preferably, a basic version of the plug in is downloadable for free. Alternatively, the user may be charged for the plug in 204. The user may be able to download an upgrade for the plug in, for which upgrade fees 208 may be charged to the user. It should be kept in mind that the plug in and/or applet could also be installed from a computer readable medium such as a floppy disk or compact disc.

Fees may be charged to the content provider based on number of downloads of content, amount of content transmitted over the network, etc. Fees may also be charged per data stream or per group of data streams up to or over a predetermined number. These would be kept track of via statistics returned to either the content provider or the technology provider. Alternatively, fees may be charged based on the size of the audience that the content provider wishes to address. Payment may also be received from a content provider for storing and hosting the vision-enabled content.

In one embodiment of the present invention, interactive advertising 210 is sent to a user. A fee may be charged to an advertiser for each time a vision-enabled advertisement is selected, such as when a user clicks on a banner advertisement 212. In another embodiment, interactive entertainment 214 is sent to the user. In yet another embodiment, interactive education 216 is sent to the user.

As an option, the plug in and encoding may be provided for free in order to collect statistics. These statistics may be made available for a fee.

Figure 3 illustrates an advertising model of the present invention corresponding to block A of Figure 2. In this model, a content publisher provides a website 300 which offers interactive advertising. Interactive advertising includes such things as encapsulated banners that begin an automatic download of the applet containing the vision-enabled content, web pages with products displayed, etc. when clicked on. Users 302 connect to the content publisher's website via a wide area network and are allowed to browse web pages of the website 300. The publisher may receive statistics on the browsing habits of the users, such as how long the user 302 was connected to the website 300 and whether the user 302 interacted with an advertisement on a web page. Further, group statistics may be collected. Also, eye tracking may be used to determine whether the user 302 looked at an advertisement. It should be noted that the advertising model is given by way of example and in no way is it intended to limit the present invention merely to advertising.

Figure 4 illustrates an exemplary process flow of the advertising model shown in Figure 3. In operation 400, a user activates an advertisement such as by clicking on a banner with a mouse. In operation 402, it is determined whether the user has the plug in and whether it is enabled. This is preferably done by the content publisher's website. If it is determined that the user does not have the plug in, it is determined whether the plug in will operate on the user's system in operation 404. This determination may be based on hardware and/or software considerations, such as whether the plug in is compatible with the user's web browser. If the plug in is not compatible with the user's computer, a standard advertisement is sent to the user's computer in operation 406 via HTML.

If the plug in is compatible with the user's computer, in operation 408 it is determined whether the user wants the plug in. If the user indicates that the user does not want the plug in, the standard advertisement is sent to the user's computer, as in operation 406. If the user indicates that the user wants the plug in, the plug in is sent to the user's station in operation 410 from either the technology provider's website or the content publisher's website. The user may then install the plug in.

If the plug in is enabled, or the user has installed the plug in, content in the form of an applet is streamed to the user in operation 412. (See the discussion of Figure 5A below for more detail on operation 412.) In operation 414, the user interacts with the applet. (See the discussion of Figure 5B below for more detail on operation 414.) When the user is finished interacting with the applet, determined in operation 416 such as by determining when the user leaves a web page or closes the applet, statistics are provided to the content producer in operation 418. In operation 420, the statistics are analyzed. These statistics can include things that are unique to this particular user, or can be combined with statistics of many users. It should be noted that the applet may be received as a single file or may be streamed to the user.

Figure 5A depicts an optional process associated with operation 412 of Figure 4 for personalizing content. First, in operation 500, a recognition of the user takes place. For example, the user may be recognized based on a cookie, an email address of the user, or user indicia. The cookie could be stored on the user's station or the advertiser's server. Alternatively, the user may be recognized based on image comparison by comparing the person image to images stored in a database. Optionally, user-entered identification indicia may be received. Such user-entered indicia could be used to allow access to an exclusive section of a website, such as one reserved for registered users of the plug in, applet, or site only.

If the user is recognized in operation 500, user information is retrieved from a database in operation 502. Such user information could include information previously input by a user, past purchases, and statistical information collected from previous browsing by the user. One example would be determining interests and/or buying habits of the user based on advertisements selected by the user in previous sessions or products previously purchased. An individualized advertisement applet is selected in operation 504 based on the user information and sent to the user in operation 506.

If the user is not recognized in operation 500, an attempt to associate the user with a group is performed in operation 508. The association may be made based on information such as the user's email address or user-input interests. Further, an

association may be imputed by country as well as from the type of site being visited: commercial, government, technical. If the user can be associated with a group, an advertisement applet is chosen in operation 510 that is targeted at the group with which the user is associated. If the user cannot be associated with a group, a standard or random advertising applet is selected in operation 512 and sent to the user as in operation 506.

Figure 5B shows exemplary processes associated with operation 414 of Figure 4. More particularly, Figure 5B illustrates what occurs at the user's station after receipt of the applet.

In a composited model of the process, an image of the user appears in the content. First, a visual image of the user taken by a camera is received in operation 520. Preferably, it is determined whether to remove a background from the image of the user in operation 522 in order to extract a person image. Removal of the background from the person image assists in the recognition of a user in the person image as well as reduces error caused by animate objects located in the background, such as a television picture. The background may also be removed to allow a portion of the person image to appear to interact with the content.

If the background of the image is to be removed, it is removed in operation 524. More information about extracting an image from its background may be found in a patent application entitled "METHOD AND APPARATUS FOR PERFORMING A CLEAN BACKGROUND SUBTRACTION" filed October 15, 1998 under Application Number 09/174,491 and which is herein incorporated by reference for all purposes.

In operation 526, body part recognition is performed on the image or extracted person image of the user to identify a head, eyes, arms, torso, etc. of the user. Further details regarding detecting body parts may be found in a patent application filed July 30, 1999 entitled "SYSTEM, METHOD AND ARTICLE OF MANUFACTURE FOR TRACKING A HEAD OF A CAMERA-GENERATED IMAGE OF A PERSON" which is incorporated herein by reference in its entirety.

In operation 528, an object such as a product for sale is composited to the image of the user by utilizing the body part recognition. For example, the user's head may be shown wearing a hat. Further details regarding compositing objects to an image of a user may be found in a patent application filed October 15, 1997 entitled "METHOD AND APPARATUS FOR MODEL-BASED COMPOSITING" under Application Number 08/951,089 and which is herein incorporated by reference in its entirety.

In operation 530, the user is given the opportunity to purchase the object with which his or her image was interacting. The purchase may be completed in operation 532. Statistics are collected in operation 534 in a manner similar to that presented above. The user is given the choice to continue or quit in operation 536. If the user wishes to continue, such as to view other objects composited to his or her image, some or all of operations 520 through 536 are repeated until the user wishes to quit. A record of some of the occurrences is offered to the user in operation 538 and created in operation 540 if the user desires one. The record could include a visual copy of the interactive session just completed, financial information if an object was purchased, and statistical information.

In an exemplary scenario, a user with the necessary plug in connects to a website with an advertisement, e.g., banner, for sunglasses. The user wishes to purchase a pair of sunglasses, but wishes to see how he or she will look wearing the sunglasses. The user clicks on the advertisement, which begins a download of vision-enabled content to the user's station. The plug in detects the camera connected to the user's station and receives an image of the user. The user's head is identified in the image of the user and may be separated from the rest of the user's body to form a person image. The user's eyes are also identified in the image of the user to determine proper placement of the sunglasses. Meanwhile, the user browses the advertisements for a pair of sunglasses to "try on." Upon selection of a pair of sunglasses, such as by pointing and clicking on a desired pair of sunglasses, the person image of the user's head is processed to composite the selected pair of sunglasses to the person image. Then, the image of the user's head is displayed

wearing the pair of sunglasses over the eyes. The user could then select different pairs of sunglasses to "try on," each of which would appear on the present person image of the user's head or on a new person image of the user's head.

Preferably, multiple images of the user turning his or her head would be captured to allow the user to manipulate the image of the head to permit viewing of a face as well as a profile for example. Two images would produce only the face and profile views. However, multiple images taken as the user turns his or her head could be used to produce the appearance of a rotating head interacting with the content. It should be kept in mind that this scenario could apply to any body part recognized in operation 526, not just the head.

Feedback may be sent to the advertiser to indicate which pair of sunglasses the user is currently looking at. Alternatively or in combination with the feedback, statistics may be sent to the advertiser upon termination of the session. Such statistics could include the amount of time the user spent looking at sunglasses, a listing of pairs of sunglasses selected, activities requested by the user, such as head rotation, etc. The statistics may then be used to create a user profile. The statistics may also be used to assist the advertiser in improving its content.

In a non-composited model of the process, the user is utilized as an input device to control the content. In other words, images of the user are used to control movement of objects in the content as well as the flow of the content. It should be kept in mind that an image of the user may still be displayed interacting with the content. First, a visual image of the user taken by a camera is received in operation 550. In operation 552, body part recognition is performed on the image of the user to identify a head, arms, or torso, etc. of the user. Preferably, multiple images of the user are received in real time via a data stream so that consecutive images may be compared to allow detection of movement.

A visual interpretation of user movement is performed in operation 554 and used to select content for display. In this way, movement of the user controls the content. In one embodiment, gesture recognition may be performed. For example, pointing up and down may be used to control scrolling of a web page, as may facing

up and down with the head. More information on gesture recognition is found in a patent application entitled METHOD AND APPARATUS FOR REAL-TIME GESTURE RECOGNITION filed October 15, 1997 under Application Number 08/951,070 and which is herein incorporated by reference in its entirety.

In another embodiment, virtual buttons may be enabled. For example, moving a hand may control movement of a cursor on the screen. Pushing the hand forward may indicate pressing a button positioned under the cursor on the screen.

In operation 556, the user is given the opportunity to purchase the object with which his or her image is interacting. The purchase may be completed in operation 558. Statistics are collected in operation 560 in a manner similar to that presented above. The user is given a choice to continue or quit in operation 562. If the user wishes to continue, operations 550 through 562 are repeated until the user wishes to quit. A record of the occurrences is offered to the user in operation 564 and created in operation 566.

Figure 6 illustrates an entertainment/educational model of the present invention corresponding to blocks B and C of Figure 2. In an entertainment model, users 600 connect to a host 602 and are allowed to request entertainment content such as audio, video, and game data. In an educational model, the users 600 are students that connect to the host 602 and request educational content such as audio and video. Content in the form of HTML and applets is sent to the users 600. Further, audio and/or images, and optionally, game data may be transmitted between the users, such as during a group game or when attending a virtual classroom. Optionally, a moderator 604 such as a referee of a game or an instructor may communicate with the host and/or the users. The moderator 604 may receive different applets than the users 600 to enable the moderator 604 to moderate a gaming or educational session.

The host 602 may receive statistics on the browsing habits of each of the users 600, such as how long a user 600 was connected to the host 602 and how long the user 600 used interactive content. Further, group statistics may be collected. The statistics may also be used during subsequent game plays to provide information about games that players particularly like playing as well as to modify a skill level of

a game for a particular player. It should be noted that the entertainment/educational model is given by way of example and in no way is it intended to limit the present invention.

Figure 7 illustrates an exemplary process flow of the entertainment/educational model shown in Figure 6. In operation 700, a user activates an entertainment or educational session by connecting to the host 602. The host may provide a listing of the vision-enabled content available to the user from which the user may choose. In operation 702, the host looks for the plug in on the user's station. Optionally, the plug in may connect to the technology provider's website to check for an upgrade in operation 704. This may occur in the background. If the plug in is current, the process continues. If the plug in is not current, the user is given the option of downloading the update in operation 706. If the user chooses to get the update, it is downloaded onto the user's station in operation 708. If the user chooses not to get the update, the process continues.

If the plug in is not found, the user is given the option to get the plug in operation 710. If the user chooses to get the plug in, it is downloaded onto the user's station, as in operation 708. If the user chooses not to get the plug in, the process is aborted in operation 712.

In operation 714, a user recognition process is performed to identify the user and/or a group to which the user belongs. This allows the host to target options towards the user or group. In the entertainment embodiment, for example, past user performance may be used to group the user with game players of similar skill. See the previous discussion with reference to Figure 5A for a description of the recognition process.

With continuing reference to Figure 7, a determination of whether the entertainment or educational activity will be interacted with by the user individually or with a group is made in operation 716. If the activity is to be performed by the user individually, options are presented in operation 718. In operation 720, the user is allowed to select the desired activity, i.e., entertainment activity or educational application, from the options presented in operation 718 and the process continues.

If the activity is to be performed by a group, options are presented in operation 722 and the user selects an activity from the options in operation 724. Based on the activity selected by the user in operation 724, an IP address for the group members is sent to the user in operation 726. This allows the members of the group to interact directly with each other without the host once the applet is received from the host, though the group activity may be performed through the host as well.

In operation 730, the applet corresponding to the selected activity is sent to the user. The user and/or group is allowed to interact with the activity in operation 732. In other words, the user plays the game, attends a virtual lecture, etc. Operation 732 is repeated until it is determined in operation 734 that the user or group has completed interacting with the activity. In operation 736, statistics are sent to the host, which may be used to create and supplement user and/or group profiles.

In an exemplary entertainment scenario, a group of users each having the proper plug in connect to an entertainment host to play a group game. Each user receives the applet associated with the game to be played from the host over a wide area network. In the game, each player is represented by an animated character. After a visual image of each of the users is obtained, a person image is recognized and body part recognition is performed on each of the images of the user to separate out a head, arms, and torso of the user, for example. The background is also separated from the person image. Then, the person image of the head of each player is composited to the animated character corresponding to that player and either the person image or data representing the animated character is distributed to each of the users. The game is played either through the host or among the players across the network. During play, each animated character bears the likeness of the associated player. Optionally, movement of a player during play is recognized and the corresponding animated character performs similar movements.

Depending on the game, interactions between the animated characters and objects appearing on the display may be required. For example, contact and collisions of the objects with the animated characters, as well as the animated characters with each other, may form part of the game, as in a game of virtual basketball. In such

case, the contact and/or collision is detected and the objects and/or animated characters are made to react accordingly. More information concerning detecting interactions between the animated characters and objects may be found in a patent application entitled "SYSTEM, METHOD AND ARTICLE OF MANUFACTURE FOR DETECTING COLLISIONS BETWEEN VIDEO IMAGES GENERATED BY A CAMERA AND AN OBJECT DEPICTED ON A DISPLAY" filed July 30, 1999 and herein incorporated by reference in its entirety.

In another exemplary entertainment scenario, a user is watching television. An applet that allows remote control of the television is enabled. The user's movements are recognized. Different movements of the user implement different commands, such as changing the volume and switching channels. Optionally, the user may be recognized upon turning the television on, and the user's favorite channel would be tuned to.

While this invention has been described in terms of several preferred embodiments, it is contemplated that alternatives, modifications, permutations, and equivalents thereof will become apparent to those skilled in the art upon a reading of the specification and study of the drawings. It is therefore intended that the true spirit and scope of the present include all such alternatives, modifications, permutations, and equivalents.